

# ORWG Meeting: Discussion and Paper Completion

Thursday 23 June 2011 @ 12:00 EDT

## Synopsis:

- Summary of discussions at the 2011 NCRI-NCI Joint Meeting (Standards and Vocabularies)
- Review of updated recommendations
- Paper refinement and completion as well as publication methods
- Project closure

# 2011 NCRI-NCI Joint Meeting

## Standards & Vocabularies Theme

### **Diverse representation**

- HL7 v2 & v3
- Significant traditional biomedical vocabularies (LOINC)
- Emerging methods, especially for Life Sciences
- Semantic Web (e.g. W3C HCLS)

### **Presentations (semi-formal, ad-hoc)**

- Survey Instruments using LOINC (and semantic profiles): Clem McDonald
- Ontology Evaluation Methods: Joanne Luciano
- BRIDG: Charlie Mead
- LS-DAM: Bob Freimuth
- HL7 Clinical Genomics: Amnon Shabo
- VIVO Project: Mike Conlin
- Imaging (radiology and pathology) integration use cases using SemWeb methods: Scott Marshall
- OBO Foundry: Alan Ruttenberg
- ORWG Project: Stuart

# 2011 NCRI-NCI Joint Meeting

## Standards & Vocabularies Theme

### Outcomes






- Meeting summary - *draft in development*
- Concise outcome/recommendations for continued work - *pending*
- White paper (including alignment with other themes) - *pending*
- Related stand-alone paper or integrated discussion on “complexity” - *pending*
- Peer-review journal (e.g. Nature) - *tentative*

All shared on Google Docs within a folder/dossier @ <http://goo.gl/Cj64w>

# 2011 NCRI-NCI Joint Meeting Standards & Vocabularies Theme

## BioSharing

[POLICIES](#) | [STANDARDS](#) | [BLOG](#) | [CONTACT](#) | [COMMUNITIES](#) | [ABOUT](#) | [LOGIN OR REGISTER](#)




“ BioSharing works at the global level to build stable linkages between journals, funders, implementing data sharing policies, and well-constituted **standardization efforts** in the biosciences domain, to expedite the communication and the production of an integrated standards-based framework for the **capture and sharing** of high-throughput genomics and functional genomic bioscience data.

We work with other organisations to

1. develop **catalogues** to centralize **bioscience data policies and reporting standards**
  - enrich these progressively by linking to other related portals and resources to serve those seeking information on systems serving or implementing the standards;
2. moderate a **communication forum** for **funders and stakeholders**
  - promote mutual support and cross-project activities to ensure the difference among the policies and standards do not impede seamless interoperability of the data. ”


[Read more about BioSharing here....](#)



### POLICIES


A catalogue of data preservation, management and sharing policies from international funding agencies and regulators.

[formats](#) | [terminologies](#) | [checklists](#)



### STANDARDS

A catalogue of reporting standards (minimum reporting guidelines, exchange formats and terminologies) and organizations that develop these.



### BLOG

Keep up to date on our activities and contribute to the discussions with our blog.

# 2011 NCRI-NCI Joint Meeting Standards & Vocabularies Theme

## VIVO




enabling national  
networking of scientists

Search

Home About Download Support Contact

## An interdisciplinary national network

Enabling **collaboration** and discovery among **scientists** across all disciplines.

[Find out how your institution can participate](#) 

The national network of scientists will facilitate the discovery of researchers and collaborators across the country. Institutions will participate in the network by installing VIVO, or by providing semantic web-compliant data to the network.




VIVO Conference 2011   
August 24-26, 2011  
Gaylord Hotel & Conference Center,  
Washington DC

[VIVO Conference 2010 materials](#)

### In the Press

May 31, 2011  
[Faculty of 1000 and VIVO: Invisible Colleges and Team Science](#)  
*Source: Issues in Science and Technology Librarianship*


May 30, 2011  
[Shared Benefits from Exposing Research Data](#)  
*Source: Griffith University, presented at IATUL*

 [More from the Press »](#)


### Events

[VIVO Implementation Fest](#)  
Jun 23 - Jun 24, 2011  
St. Louis, MO

[ALA 2011 Annual Conference](#)  
Jun 23 - Jun 29, 2011  
New Orleans, LA

 [All Events »](#)

### Blog

 [Keynote Speakers - 2011 VIVO conference](#)


We are pleased to announce [Keynote Speakers](#) for the [Second Annual VIVO Conference](#): Dr. Simon Liu, Director of the National Agricultural Library (NAL) and Dr. Barend Mons,...


 [Read the Blog »](#)



# 2011 NCRI-NCI Joint Meeting Standards & Vocabularies Theme

## BioSiteMaps | Biomedical Resource Ontology

**BIO SITEMAPS**



Home

**Biositemaps**

Breaking News: The [Resource Discovery System](#) now includes biositemap resources from all registered resource programs!

**The Biositemaps Working Group of the NIH Roadmap National Centers of Biomedical Computing (NCBC) ([www.ncbcs.org](http://www.ncbcs.org))** has developed technologies to address (i) locating, (ii) querying, (iii) composing or combining, and (iv) mining biomedical resources. Each site which intends to contribute to the inventory instantiates a file on its Internet site "**biositemap.rdf**" which conforms to a defined RDF schema and uses concepts from the Biomedical Resource Ontology to describe the resources.

**What is a Biositemap?** Biositemaps represent a mechanism for computational biologists and bio-informaticians to openly broadcast and retrieve meta-data about biomedical resources. All institutions with an interest in biomedical research can publish a biositemap.rdf file on their Internet site. Each biositemap.rdf file is simply a list of controlled metadata about resources (data, software, tools, material, and services) that your organization uses or believes are important to biomedical research. The key enabling technologies are the [Information Model \(IM\)](#) which is the list of metadata fields about each resource (resource\_name, description, contact\_person, resource\_type,...) and the [Biomedical Resource Ontology \(BRO\)](#) which is a controlled terminology for the 'resource\_type', 'area of research', and 'activity' and which are used to improve the sensitivity and specificity of web searches. [\[more...\]](#)

**What is the Biomedical Resource Ontology (BRO)?** A key enabling technology for Biositemaps is the Biomedical Resource Ontology (BRO) which is a controlled terminology for the 'resource\_type', 'area of research', and 'activity' and which are used to improve the sensitivity and specificity of web searches. This is under development by a number of NIH-funded researchers who have a combined interest in classification of biomedical resources. The publication site for [BRO is the BioPortal](#).

**QuickStart for Biositemaps**

1. Use the [Biositemaps Editor](#) to fill in the information about the resources on your site and generate a biositemap.rdf file. There are also customized editors for [CTSA](#) and [NCBC](#). Important: **Internet Explorer** users must first [configure](#) their browser. [\[help...\]](#)
2. Your biositemap.rdf file will need to be **deployed** to your local web server prior to being published [\[help...\]](#)
3. Use the [Biositemap Registration](#) to publish your biositemap after it has been deployed [\[help...\]](#)  
*Biositemaps need only be published once, and may be freely updated after their initial publication.*
4. You can query information about resources that you or others have published using the [biositemaps search tool](#).

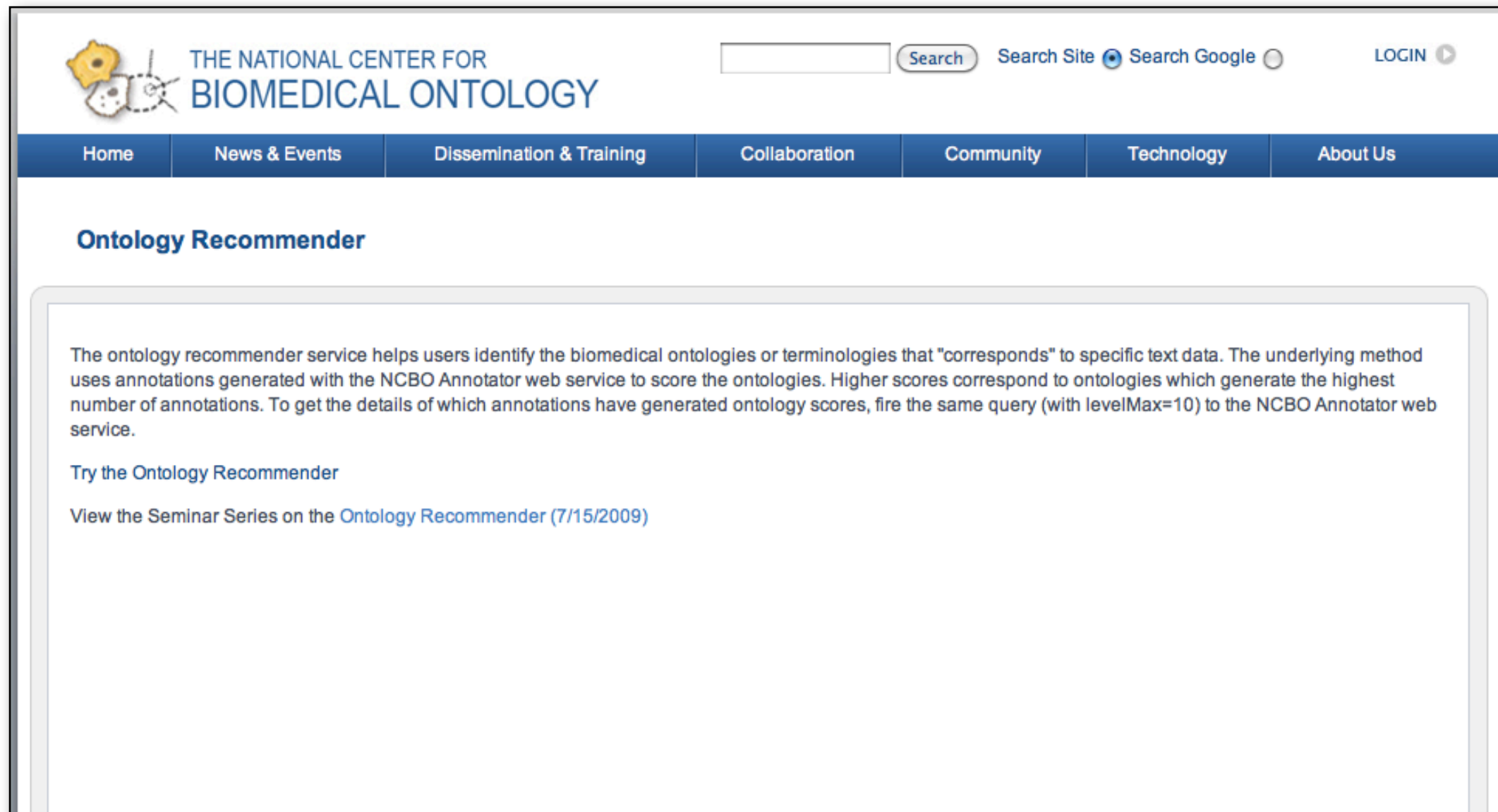
**How can researchers author and consume Biositemaps and the BRO?**

- The [Biositemaps Editor](#) provides an authoring web interface to fill in the information about the resources on your site and generate a biositemap.rdf file. There are also customized editors for [CTSA](#) and [NCBC](#). Important: **Internet Explorer** users must first [configure](#) their browser.
- The [Biositemap Search](#) provides a web interface for simple or complex queries against all resources defined in all published biositemaps across the internet.



# 2011 NCRI-NCI Joint Meeting Standards & Vocabularies Theme

## Ontology Recommender



The screenshot shows the top of the NCBO website. The header includes the NCBO logo (a stylized orange and white flower-like shape) and the text "THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY". To the right of the logo is a search bar with a "Search" button, and links for "Search Site" and "Search Google". Further right is a "LOGIN" link with a play button icon. Below the header is a navigation bar with seven tabs: "Home", "News & Events", "Dissemination & Training", "Collaboration", "Community", "Technology", and "About Us". The main content area is titled "Ontology Recommender" in blue. Below this title is a text box containing the following text: "The ontology recommender service helps users identify the biomedical ontologies or terminologies that 'corresponds' to specific text data. The underlying method uses annotations generated with the NCBO Annotator web service to score the ontologies. Higher scores correspond to ontologies which generate the highest number of annotations. To get the details of which annotations have generated ontology scores, fire the same query (with levelMax=10) to the NCBO Annotator web service." Below this text are two links: "Try the Ontology Recommender" and "View the Seminar Series on the Ontology Recommender (7/15/2009)".

THE NATIONAL CENTER FOR  
BIOMEDICAL ONTOLOGY

Search Search Site Search Google LOGIN

Home News & Events Dissemination & Training Collaboration Community Technology About Us

### Ontology Recommender

The ontology recommender service helps users identify the biomedical ontologies or terminologies that "corresponds" to specific text data. The underlying method uses annotations generated with the NCBO Annotator web service to score the ontologies. Higher scores correspond to ontologies which generate the highest number of annotations. To get the details of which annotations have generated ontology scores, fire the same query (with levelMax=10) to the NCBO Annotator web service.

[Try the Ontology Recommender](#)

View the Seminar Series on the [Ontology Recommender \(7/15/2009\)](#)

# 2011 NCRI-NCI Joint Meeting

## Standards & Vocabularies Theme

### **Cross Community Pollination**

- Language (e.g. “metadata”)
- Intellectual Property Issues
- Use Cases require (may) different solutions: patient care vs. life science research

### **Profiling**

- Language (e.g. “metadata”)
- Intellectual Property Issues
- Use Cases require (may) different solutions: patient care vs. life science research
- User centered design (UCD)
- Problems: incentives, metrics, specific ontology community size, absolute community size, web-of-trust, community fragmentation
- Ontology best practices and quality control: Unit testing, reasoners, education and training



# 2011 NCRI-NCI Joint Meeting

## Standards & Vocabularies Theme

### Framework - Methods and Metrics

- Evaluation metrics (Joanne Luciano - see [SlideShare presentation](#))
- Search behavior - disparities (browse for whole ontologies task or domain, browse terms in multiple ontologies)
- Reconciling universal objectives for metrics (text mining vs. reasoning)
- Critique of utility of existing metadata available in repositories (e.g. BioPortal)
- Potential for automatic evaluation. Approach similar to rigors of software development (e.g. unit testing)
- High priority metrics: Domain or usage or other coverage, use of axioms to determine how to use a specific term, classification methods
- Incentives to crowd source reviews: People require recognition and ability to author their own review. Presence and reputation is under-appreciated

# 2011 NCRI-NCI Joint Meeting

## Standards & Vocabularies Theme

### **Connecting different models or moving between different formalizations**

- Evaluation metrics (Joanne Luciano - see [SlideShare presentation](#))
- Search behavior - disparities (browse for whole ontologies task or domain, browse terms in multiple ontologies)
- Reconciling universal objectives for metrics (text mining vs. reasoning)
- Critique of utility of existing metadata available in repositories (e.g. BioPortal)
- Potential for automatic evaluation. Approach similar to rigors of software development (e.g. unit testing)
- High priority metrics: Domain or usage or other coverage, use of axioms to determine how to use a specific term, classification methods
- Incentives to crowd source reviews: People require recognition and ability to author their own review. Presence and reputation is under-appreciated

# 2011 NCRI-NCI Joint Meeting

## Standards & Vocabularies Theme

### Scenarios

Adoption, traditional and novel use of “ontologies” may be helped by better publishing or sharing highly relevant integration/interoperability scenarios. This is expected to improve collaboration between those who need help with interesting solutions to novel problems (e.g. researchers and the semantic web community)

Possible exemplars:

- UK Database of Uncertainties About the Effects of Treatments (DUET)  
“A resource to make uncertainties explicit and to help prioritise new research”
- KickStarter (funding platform for creative projects)

# ORWG Paper: Updates (some)

Regarding explicit ratings, NCBO is likely the best single location to serve as a repository of ratings and rankings by the community. However, as discussed earlier, it is recommended that some changes to the NCBO site be made to create more of a community neighborhood gathering place, and summary of terminology ratings, linking to the actual ratings on each terminology. To create an initial critical mass, NCBO should consider inviting a major user of each ontology to create an initial rating entry. Also, they might consider adding the “Did you find this useful” feature so prevalent on other ratings sites

Related to 2 above, we recommend bringing all terminology metadata content to the surface, (example Ohloh.net, Sourceforge, Amazon are good examples). BioPortals community features have been buried too deeply, but interim changes in the Bioportal user interface in May 2011 release have been made to help bring this information closer to the surface

Description of rights in the current models are fairly general \- for instance Dublin Core uses "accessRights" and OMV uses "license model". Additional information about (or link to) attribution, reuse and distribution should be included by submitters, as this is needed by users.&nbsp; Although this is not in the current core model recommendation, guidance should be provided to contributors to provide this information in a note.

# ORWG Paper Completion

- End of June
- Relatively major reworking (condensation) with references to notes and focus on recommendations.
- Publication formats
- Update meeting with NCRI (Stuart Bell and Sherri de Coronado) week of 13 June 2011